sai®

Simeone Associates, Inc.

Estimating the Size of an Illicit Drug Using Population

Project: RAMONA
Deliverable: A.1 – A.4
October 7, 2000

Simeone Associates, Inc
220 Lancaster Street
Albany NY 12210
518.428.9323 (Voice)
518.449.8265 (Fax)

Executive Summary
Estimating the Size of an Illicit Drug Using Population

This report summarizes work that has been completed on the modeling component of the Random Access Monitoring Of Narcotics Abusers (RAMONA) project. The contract called for a review of the methodology used in the initial Cook County study, and an assessment of whether some simpler model specification might be employed in the current design. A simulated system of drug use and treatment-seeking behavior was to be developed, and the ability of the model to recover the parameters of the system was to be tested. Finally, the model was to be applied to the original Cook County data to determine if any significant differences in results exist. Taken together, these activities constitute contract tasks A.1 through A.4. This report is the deliverable product associated with those tasks.

We began with a model of the drug use career that is based on a simple exponential process. The model operates within a competing risks framework. It assumes that an individual who initiates a spell of drug use has at any moment some non-zero probability of entering treatment or ending his spell. The model also assumes that data on drug use careers are collected as drug users present for treatment, and that as such the sample is truncated.

We then developed a system to simulate drug use and treatment-seeking behavior. The mechanics of the system mirror the assumptions associated with our model. Cohorts of individuals begin spells of drug use at regular intervals, and at any time during these spells they are at risk of entering treatment or ending their spell. The parameters of the process that generates drug use careers can be manipulated, including the rate of spell incidence, the mean spell duration, and the mean duration between treatment events. Taken together, these factors determine when the simulated population reaches equilibrium. Alternative sampling scenarios can be implemented as well. The system thus allows the duration of the sampling window and the length of the retrospective period for the Life History Interview (LHI) to be varied across trials.

We next examined the ability of the model to recover the parameters of the simulation in a series of scenarios that varied the rate of spell incidence, the mean spell duration, and the mean duration between spells. We found a very high degree of concordance between model estimates and simulation parameters, and the results of a typical scenario are provided in a later section of this report. These findings tell us that if the simulation provides a reasonable description of the process by which people use drugs and seek treatment, then the model will provide a correspondingly reasonable estimate of the size of the drug using population.

Finally, we applied the model to the original Cook County data. This required that we develop a number of routines designed to read the LHI files and assemble careers for individuals who had been sampled in that study. The routines allow alternative thresholds to be set for the drugs that are of interest to us. For the purposes of this research, we adopted the same thresholds that were used in the earlier work. Also, consistent with the intent of this task, we based our estimates only on individuals whose career data were collected at treatment programs. The principal objective was to estimate the rate at which drug users generate treatment admission events. The Cook County study produced a rate of about 0.15 per year, and our model produces a rate of about 0.14. The estimates thus appear to be reasonably close, particularly since no explanatory variables were included in our analysis. Taken together with the results of the simulation study, this outcome is very encouraging.

Over the course of the next year, we will examine various extensions of the model that are likely to be relevant to the RAMONA estimation procedure. This will involve following the protocol outlined above. We will develop a particular model specification, test the ability of this model to recover the parameters of some simulated system of drug use and treatment seeking behavior, and then (where appropriate) apply the model to the Cook County data.

Explanatory variables will be introduced to allow treatment seeking behavior to vary across individuals. The impact on the estimates of left censoring will be examined as well. We know that the LHI must have a finite retrospective period, and as a practical matter, we would like this to be as short as possible. In addition, the assumption that the rate of transition from one event to another remains constant will be relaxed. Allowing the rate of transition to accelerate over the course of the spell, corresponding to a reinforcement effect, further generalizes the model. The advantages of bootstrapping versus analytic techniques in developing standard errors for our estimates will also be assessed.

Estimating the Size of an Illicit Drug Using Population

Introduction

This report describes a method for estimating the size of an illicit drug using population. The approach forms the basis for what is intended to become a series of national estimates that will be developed under the Random Access Monitoring Of Narcotics Abusers (RAMONA) program. It builds upon the results of a feasibility study, funded by the Office of National Drug Control Policy (ONDCP), and conducted in Cook County, Illinois.[1]

The method compensates for two sources of underreporting commonly associated with drug use surveys: That individuals are reluctant to acknowledge illicit behavior within the context of a research interview; and that individuals engaged in illicit behavior may be difficult to find in substantial numbers. It deals with these problems by collecting information on the careers of admitted drug users who appear at certain locations, such as drug treatment programs, in large numbers.[2] A Life History Interview (LHI) is used for this purpose, allowing respondents to report on spells of drug use, and on the timing of various events that may occur during these spells.[3] Data from the LHI are used to build a mathematical model of the drug use career.[4] The model allows the rate at which drug users generate events of various kinds to be determined. Estimation then becomes a matter of dividing some known number of events by the rate at which individuals generate such events.[5]

The Cook County study produced estimates based upon three sources of information, involving LHIs collected at booking facilities, treatment programs, and homeless shelters. Generally, the results indicated that there were about three times as many users of heroin and cocaine in that area of the country as one would conclude if one were to rely upon more conventional sources of information, such as the National Household Survey on Drug Abuse (NHSDA).

[1] See Simeone, *et al.* (1997).

[2] A number of authors have reported on the discrepancies that exist between self report data and laboratory test results for the same individuals. Please see Harrison (1990); Hubbard, Marsden and Rachel (1989); McNagny and Parker (1992); Mieczkowski (1990, 1992); Toborg, *et al.* (1989); and Wish, O'Neil, and Baldau (1990).

[3] The LHI is a calendar instrument used to collect data on past events. For a detailed discussion, please see Adams and Henley (1977); Anglin, Hser, and Chou (1993); Freedman *et al.* (1998); and Sobell, Sobell, and Riely (1988).

[4] For a discussion related to the career construct, please see Anglin and Speckart (1986); Ball, Shaefer, and Nurco (1983); Becker (1963); Dai (1937); Hanlon, *et al.* (1990); Maddux and Desmond (1986); Nurco, *et al.* (1988); Simpson and Marsh (1986); and Waldorf (1973).

[5] Among sampling statisticians, the general technique is called ratio estimation (please see Cochran (1977)). The capture-recapture models that have been broadly applied in the field of drug use prevalence estimation are special cases of ratio estimation. Please see Brecht and Wickins (1993); Hser, Anglin, and Homer (1992); Jolly (1965); Seber (1965,1973); Simeone, Nottingham, and Holland (1993); Simeone, Rhodes, and Hunt (1995); Simeone, *et al.* (1997); Wickins (1993); and Woodward, Bonnett, and Brecht (1985).

The study results also suggested that it would be possible to develop estimates of the number of drug users with a more simple model specification, and based upon data collected at drug treatment programs alone.

This paper examines these issues. We begin with the development of a general model to use in estimation. The model is then tested in simulation, the objective being to recover the parameters of drug use and treatment-seeking activity as they exist within a contrived reality. We then apply the model to the Cook County data, for the purpose of demonstrating that comparable results can be obtained using LHIs collected at drug treatment programs alone. Finally, we discuss directions for future research to be pursued under RAMONA.

An Approach to Measurement

The approach to measurement presented below allows us to view drug use and treatment-seeking behavior as parts of an integrated data generating process. From this perspective, the careers of individuals who seek treatment during a particular period of time provide a window into the entire population of drug users, including those who do not seek treatment during that period of time.

Let N be the size of a stable illicit drug using population. Suppose all users who enter treatment over the course of a year are sampled and that individual decisions to enter treatment are made independently. Define π_i as the probability that the i^{th} illicit drug user was sampled. Then n, the number of individuals sampled, can be specified as follows,

$$n = \sum_{i=1}^{N} \delta_i \tag{1}$$

where

$$\begin{aligned} \delta_i &= \begin{cases} 1 & \textit{with probability } \pi_i \\ 0 & \textit{with probability } (1-\pi_i) \end{cases} \\ E(\delta_i) &= \pi_i \\ Var(\delta_i) &= \pi_i(1-\pi_i) \end{aligned} \tag{2}$$

Then

$$\begin{aligned} E(n) &= \sum_{i=1}^{N} \pi_i \\ &= N\pi \end{aligned} \tag{3}$$

$$\begin{aligned} Var(n) &= \sum_{i=1}^{N} \pi_i(1-\pi_i) \\ &= N\bar{\pi}(1-\bar{\pi}) - \sum_{i=1}^{N} (\pi_i - \pi)^2 \\ &= N\left[\pi\,(1-\bar{\pi}) - Var(\pi_i)\right] \end{aligned} \tag{4}$$

An estimate of $\bar{\pi}$ from the sample will yield an estimate of the entire drug using population. Let $\widehat{\alpha}$ be an estimator for $1/\bar{\pi}$, with variance, $Var(\widehat{\alpha}|n)$. Then

$$\hat{N} = n/\pi \tag{5}$$

with

$$\begin{aligned} E(\hat{N}) &= N \\ Var(\hat{N}) &= N\left[\frac{1-\pi}{\pi} - \frac{Var(\pi_i)}{\bar{\pi}^2}\right] + n^2 Var(\widehat{\alpha}|n) \end{aligned} \tag{6}$$

The remainder of this paper pertains to the determination of an estimator for $\bar{\pi}$, the average rate at which drug users generate treatment events over the course of a year. Since this average must be taken over all drug users, its estimation is problematic in that those who do not enter treatment at any point during the study period are unobservable. This implies that the only available sample from which we can obtain our estimate is not random but truncated.

Our solution to this problem is based on two assumptions. First, we assume that the process which determines when an individual drug user chooses to seek treatment possesses both a deterministic and a stochastic component. The deterministic component may be related to the personal characteristics of the user, such as gender, age, or race. We refer to this as the user profile. Second, we assume that for users with a given profile, the only differences between those who seek treatment during a particular period of time and those who do not are attributable to the stochastic component, or chance. If these assumptions are true, then we can learn much of what we need to know about the event generation process from the histories of individuals who seek treatment during the observation period.

There may be some drug users who do not seek treatment under any circumstances. This would indicate a fundamental split in the population between those who are treatment susceptible and those who are not, leaving us with the capability of estimating only the number of drug users who have a non-zero probability of ever entering treatment. While this is a potential limitation of the approach, the problem may not be of great significance.

The Cook County research proceeded from the assumption that as the number of site types included in the sample increased, the size of the population excluded from coverage would decrease. The reasoning was simple. While there may be some drug users who have zero probability of seeking treatment, there are relatively few who have zero probabilities of both seeking treatment and being arrested (and by extension, fewer still who have zero probabilities of seeking treatment, being arrested, and experiencing some third type of event).

The results show that comparable estimates of the size of the illicit drug using population were derived from the treatment admission and arrest samples, suggesting that the marginal benefit associated with including more than one site type may be low. We conclude from this that the approach described here is likely to provide suitable coverage of the population that is the intended object of study for RAMONA.

The Model

The rate at which an individual generates treatment events can be said to summarize the durations between events.[6] This is due to the reciprocal relationship between the rate at which events occur and the average length of the durations. A high rate implies short intervals between events; the opposite is also true. Therefore, we can obtain the rate by estimating the parameters of the process that generates treatment events or, more precisely, the durations between events.

In general, the durations between events may not be distributed identically across a sequence of events. The duration distribution may also vary depending on the user profile. For simplicity of exposition, we will abstract from all distinguishing characteristics other than the event number (we will abstract from that distinction as well in an example below). Define

$$F_k(t) = \Pr\{t_k \leq t\} = \textit{ the probability that the } k^{th} \textit{ treatment event during a continuous spell of drug use occurs within } t \textit{ periods after the } (k-1)^{th} \textit{ event},$$

and

$$f_k(t) = \Pr\{t_k = t\} = \textit{ the probability that the } k^{th} \textit{ treatment event during a continuous spell of drug use occurs exactly } t \textit{ periods after the } (k-1)^{th} \textit{ event}.$$

These distributions must incorporate the possibility that the user's spell may end before the k^{th} event occurs. We therefore conceptualize drug use behavior as a simple competing risks model.[7] At each point during the spell, the probability that the user will seek treatment competes with the probability that he will stop using drugs altogether. To cover both of these possibilities, we define

$$F_{1k}(t) = \Pr\{t_{1k} \leq t\} = \textit{ the probability that the } k^{th} \textit{ treatment event occurs within } t \textit{ periods after the } (k-1)^{th} \textit{ event},$$

[6] Consider the relationship between a Poisson process and the exponential distribution.

[7] Please see Basu and Ghosh (1995).

and

$$F_{2k}(t) = \Pr\{t_{2k} \leq t\} = \textit{the probability that the individual ends his spell within } t \textit{ periods after the } (k-1)^{th} \textit{ event.}$$

Following convention, we define $\overline{F}(t)$ to be the complement of $F(t)$, or $1 - F(t)$. Then we can specify $F_k(t)$, the probability that the k^{th} treatment event occurs within t periods after the $(k-1)^{th}$ event, given that the spell does not end before time t, as[8]

$$F_k(t) = 1 - \overline{F}_{1k}(t)\overline{F}_{2k}(t) \tag{7}$$

We can define the density for the realization of K events with specified intervals within the total interval $[Y_0, Y_T]$ as

$$h(K,\ t_1, \cdots,\ t_K) = f_1(t_1)\ f_2(t_2)\ \cdots\ f_K(t_K) \tag{8}$$

where the sum of the t_k's is T.

Our goal is to estimate the parameters of the distribution of durations between events, based on a sample of drug users obtained at treatment sites. The latter is not a random sample of drug users in that it is characterized by truncation. Hence, estimation based on (8) alone will produce biased results. Such an estimate would overstate the rate, since at least one treatment event is required to enter the sample. To eliminate this bias, we must condition on the likelihood of entering the sample, that is, the probability that an individual whose spell of drug use began at Y_0 seeks treatment sometime during the sampling period. Define

$$[Y_A,\ Y_B] = \textit{the sampling period;}$$

$$T_A = Y_A - Y_0;\ \textit{the interval between the start of the spell and the sampling period; and}$$

$$R = Y_B - Y_A;\ \textit{the length of the sampling period.}$$

We want to know the probability that the user experiences her k^{th} event within a period of length t. For example, suppose k is equal to 2. Then the probability that the user experiences her second event within a period of length t is the probability that the sum of the first two durations is less than or equal to t.

[8]It has been noted by several authors that depending on the sampling, the parameters of both F_1 and F_2 may not both be identifiable. Please see Basu and Ghosh (1995).

We can determine this by holding the first duration fixed at t, finding the probability that the second duration is less than or equal to $t - t_1$, and then integrating over all values of t_1 between 0 and t. This gives us

$$\Pr\{t_1 + t_2 \leq t\} = \int_0^t f_1(t_1) F_2(t - t_1) dt_1 \tag{9}$$

Let $C_k(t)$ represent the probability that the sum of the durations between the user's first k events is less than or equal to t. Then using (9) to determine $\Pr\{t_1 + t_2 + \cdots + t_K \leq t\}$ gives

$$C_k(t) = \int_0^t \int_0^{t - s_1} \int_0^{t - s_1 - s_2} \cdots \int_0^{t - s_1 - s_2 - \cdots - s_{K-2}} f_1(s_1)\, f_2(s_2) \cdots f_{k-1}(s_{k-1})$$
$$F_k(t - s_1 - s_2 - \cdots - s_{k-1}) ds_{k-1} \cdots ds_2 ds_1 \tag{10}$$

If the rate of initiation into drug use remains constant, with spell start dates assumed to be uniformly distributed, then the probability of the user having an event—*any event*—during a drug use spell of length T is

$$P(T) = \sum_{k=1}^{\infty} C_k(T)$$

By extension, the probability of the user having an event during the sampling period is

$$P(R) = \sum_{k=1}^{\infty} C_k(T_A + R) - \sum_{k=1}^{\infty} C_k(T_A) \tag{11}$$

Then the likelihood of a particular observation occurring is

$$L(K,\ t_1, \cdots,\ t_K | Y_0) = \frac{f_1(t_1)\ f_2(t_2) \cdots f_K(t_K)}{\sum_{k=1}^{\infty} C_k(T_A + R) - \sum_{k=1}^{\infty} C_k(T_A)} \tag{12}$$

Testing the Model Through Simulation

We test the effectiveness of the model developed in the previous section by applying it to a set of data simulated from a known distribution with one parameter. For purposes of exposition, we choose the exponential distribution to describe the length of the durations between treatment events. We assume further that this distribution remains constant across event number – truly the simplest case. Hence, for $k = 1, 2, ..., K$

$$f(t) = \alpha e^{-\alpha t} \tag{13}$$

Depending on the problem at hand, the parameter a is alternately referred to as the transition rate, failure rate, rate of decay, or hazard rate. In the present case, the simple exponential distribution possesses great heuristic value, since as the reciprocal of the mean of the distribution, a represents precisely what we are looking for—the rate at which drug users, on average, generate treatment events. For example, for $a = 0.025$, the mean duration between events is 40 months, implying that the typical drug user generates an average of 0.025 treatment events per month.

The data generating process replicated by the simulation can be summarized as follows.[9] Time units represent months. Each month, cohorts of identical size initiate spells of drug use. Consistent with the competing risks model, the behavior of each individual user is fully circumscribed by two independent exponentially distributed random variables: (1) the duration until the next treatment event, and (2) the duration until the end of the spell. Upon initiation, realizations from each distribution are drawn. If the realization of the duration until the first event is less than that of the duration until the end of spell, then an event is recorded for that observation and a new interval of drug use within the spell begins. This process continues until a pair of draws produces a duration until end of spell that is less than the duration until the next event. Once the spell ends, no more activity for that observation is recorded.

In this example, we set the transition rate between events, a_1, equal to 0.030, implying a mean duration between events of $33.\overline{3}$ months; and we set the transition rate to the end of spell, a_2, equal to 0.020, implying a mean duration of 50.0 months. Because the end of a spell of drug use can not be observed, a_2 can not be estimated (see footnote 7). Once the population of drug users reaches equilibrium, a sample is drawn comprising all users who generate an event during an arbitrarily chosen 12-month period. Their careers are censored from the right at the point at which they enter the sample. We seek to evaluate from the sample the rate at which the simulated users generate events.

Because we know exactly how many drug users there are in the simulated population, we can obtain π directly. When 20 new spells are initiated each month, the system reaches equilibrium with a population of 1,301. These individuals generate 384 events during an arbitrarily chosen 12-month period. This yields a rate of 0.2952 events per drug user, per year. However, the parameters of our simulation pertain to a month of drug use. We can convert the annualized rate to a monthly rate by dividing by the average number of months that the drug users are active during the 12-month period, which is in this case 9.7555. This yields a rate of 0.0302 events per month, almost precisely the value of a_1.

[9] A detailed description of the simulation model used to generate the data is available from SAI.

Our simulated data collection procedure produces a sample of 327 users. Based on equation (12), the log likelihood function is constructed as

$$\ell(K,\ t_1,\cdots,\ t_K|Y_0) = \sum_{i=1}^{327}\left\{\sum_{k=1}^{K^i}\ln f(t_k^i) - \ln\left[\sum_{k=1}^{\infty}C_k(T_A^i+R) - \sum_{k=1}^{\infty}C_k(T_A^i)\right]\right\} \tag{14}$$

where $K,\ t_1,\cdots,\ t_K$, and Y_0 are now 327×1 vectors. We can substitute the simple exponential distribution (13) into (14) and utilize the fact that the sum of k independent exponentially distributed random variables with transition rate a has a gamma distribution with parameters a and k. This yields

$$\ell(K,\ t_1,\cdots,\ t_K|Y_0) = \sum_{i=1}^{327}\left\{K^i\ln a - a\sum_{k=1}^{K^i}t_k^i - \ln\sum_{k=1}^{\infty}\left[C_k(T_A^i+R) - C_k(T_A^i)\right]\right\} \tag{15}$$

where

$$C_k(t) = \frac{a^k}{(k-1)!}\int_0^t \tau^{k-1}e^{-a\tau}d\tau$$

Equation (15) can now be maximized with respect to a using standard maximum likelihood techniques; the value obtained is 0.0291.[10]

An Example: The Cook County Data

Estimating the parameters of equation (12) necessarily involves placing some restrictions on the event generating process. For the purpose of applying our technique to the Cook County data, we maintain the simplicity of the exponential model described above. Our goal is to determine how the result based on this model compares to that obtained in the Cook County study. The findings reported there indicate that drug users generate 0.15 treatment events per year of drug use.[11] Because our interest is principally in the rate at which drug users seek treatment, and because we want to test the model developed in the previous section, we restrict the sample to only those LHIs collected at treatment sites. This leaves us with 376 cases. Estimating (15) for this sample produces a value for a of 0.0117. This is the average number of treatment events generated per drug user, per month of drug use. The average number of treatment events generated per year of drug use is therefore 0.14.[12] This is very close to the estimate reported in the earlier research.

[10]Some of the simulated drug users experience additional treatment events within the 12-month window. However, because the cases are right-censored at their first treatment in the window, these additional events are not included in the analysis. This accounts for a small downward bias in the estimate. Additionally, to reduce computation time, we limit the sum in the denominator of (15) to 10. Because the probability of experiencing more than 10 events is very small, the impact on the solution is barely detectable.

[11]Please see Simeone, *et al.* (1997).

[12]We note that the present analysis excludes explanatory variables that were used in the Cook County study, as these data were not readily available. Therefore, our results may be affected by omitted variable bias.

Extensions

Over the course of the next year, we will examine various extensions of the model that are likely to be relevant to the RAMONA estimation procedure. This will involve following the same protocol outlined in the preceding sections. We will develop a particular model specification, test the ability of the model to recover the parameters of some simulated system of drug use and treatment seeking behavior, and then (where appropriate) apply the model to the Cook County data.

Explanatory Variables

A critical extension of the model presented above is the ability to control for elements of the user profile that are most likely to produce systematic variability in the rate at which individuals seek treatment. Previous research suggests that such characteristics would include ethnicity, gender, age, and primary drug of abuse. Preserving the exponential distribution for the durations between events motivates an exponential regression model where the error term $t_i/g(x_i)$ is distributed exponentially with a transition rate of 1. Thus

$$t_i/g(x_i) \backsim e(1) \tag{16}$$

where $g(x_i)$ is the expected value of the duration between events, and x_i is a $1 \times p$ vector of covariates.[13] It is convenient to develop the model in terms of the log-durations, $y_i = \ln(t_i)$. The model can be written as

$$y_i = x_i\beta + \varepsilon_i \tag{17}$$

where β is a $p \times 1$ vector of regression coefficients and the ε_i are assumed independent. Equation (8) can now be replaced by

$$\begin{aligned} h(K,\ y_1, \cdots,\ y_K, \beta | X) = &\ \prod_{i=1}^{n} \prod_{k=1}^{K} \exp\left[(y_{ik} - x_i\beta) - \exp(y_{ik} - x_i\beta)\right] \\ = &\ \exp\left[\sum_{i=1}^{n} \sum_{k=1}^{K} (y_{ik} - x_i\beta) \right. \\ &\left. - \sum_{i=1}^{n} \sum_{k=1}^{K} \exp(y_{ik} - x_i\beta)\right] \end{aligned} \tag{18}$$

where X is the $n \times p$ design matrix and x_i is the i^{th} row of X.

Left Censoring

When the LHI does not extend back to the beginning of the user's current spell of drug use, the spell is said to be left-censored. This produces an upward bias in the rate of treatment seeking if not taken into account. We can eliminate this bias by replacing $f_1(t_1)$ with $1 - F_1(t_1)$, representing the fact

[13] Please see Viveros (1995) and Greene (1993).

that the only information we have on the first duration is that it is at least of length t_1.

Reinforcement Effects

Another extension relates to the relaxation of the assumption that the distribution of durations is constant across the user's sequence of treatment events. A simple choice for allowing $f_k(t)$ to vary is to assume that the mean duration rises or falls with the number of events already experienced, implying the presence of contagion.[14] Following Coleman, equation (13) can be generalized to yield

$$f_k(t) = [a + (k-1)b]\, e^{-[a+(k-1)b]t} \tag{19}$$

Although the simple exponential distribution provided an adequate fit to the Cook County data, we should be prepared to consider alternative functional forms if appropriate.

Standard Errors

Within the maximum likelihood framework, the computation of standard errors is based on the asymptotic covariance matrix for the maximum likelihood estimator. In general, the latter matrix is the inverse of the information matrix, $I(\theta)$, which is the negative expected value of the matrix of second derivatives of the log likelihood function with respect to θ. Thus

$$[I(\theta)]^{-1} = \left\{ -E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta\partial\theta'} \right] \right\}^{-1} \tag{20}$$

Even in the single parameter case, with $\ln L$ defined in equation (15) above, this is an extremely cumbersome computation. Moreover, its finite sample properties are unknown. For example, in order for the asymptotic theory of maximum likelihood to be applicable, the sample size must be large. However, it is difficult to know exactly how large. In cases similar to the present one, a number of authors have recommended a technique known as bootstrapping for the purpose of deriving standard errors for the parameter estimates.[15] Bootstrapping permits the estimation of standard errors based purely on the finite sample behavior of the available data.

[14] For a definition and examples of contagious processes, please see Coleman (1964) and Allison (1980).

[15] Please see McLachlan (1995) and Zacks (1995).

References

Adams, L. and J. Henley (1977). "Measuring various drug use dimensions using a calendar method," *International Journal of the Addictions*, 12: 423-27.

Allison, P.D. (1980). "Estimation and Testing for a Markov Model of Reinforcement," *Sociological Methods & Research*, 8(4): 434-453.

Anglin, M., Y. Hser, and C. Chou (1993). "Reliability and validity of retrospective behavioral self-report by narcotics addicts," *Evaluation Review*, 17: 91-108.

Anglin, M., and G. Speckart (1986). "Narcotics use, property crime and dealing: Structural dynamics across the addiction career," *Journal of Quantitative Criminology*, 2: 355-375.

Ball, J.C., J.W. Shaefer, and D.N. Nurco (1983). "The day-to-day criminality of heroin addicts in Baltimore-A study in the continuity of offense rates." *Drug and Alcohol Dependence*, 12: 119-142.

Basu, A.P. and J.K. Ghosh (1995), "Competing risks theory and identifiability problems," in N. Baladrishnan and A.P. Basu (eds.) *The Exponential Distribution: Theory, Methods and Applications*. Amsterdam: Gordon and Breach Publishers.

Becker, H. (1963). *Outsiders: Studies in the Sociology of Deviance*. Glencoe: Free Press.

Brecht, M. L., and T.D. Wickins (1993). "Application of multiple-capture methods for estimating drug use prevalence," *Journal of Drug Issues*, 23: 229-250.

Coleman, J.S. (1964). *Introduction to Mathematical Sociology*. London: The Free Press of Glencoe.

Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley and Sons, Inc.

Dai, B. (1937). *Opiate Addiction in Chicago*. Shanghai: Commercial Press.

Freedman, D., A. Thornton, D. Camburn, D. Alwin, and L. Young-DeMarco (1988). "The life history calendar: A technique for collecting retrospective data," *Sociological Methodology* 18: 37-68.

Greene, W.H. (1993). *Econometric Analysis*. New York: Macmillan Publishing Co.

Hanlon, T., D. Nurco, T. Kinlock, and K. Duszynski (1990). "Trends in criminal activity and drug use over an addiction career," *American Journal of Drug and Alcohol Abuse*, 16: 223-238.

Harrison, L. (1990). *The Validity of Self-Reported Drug Use Among Arrestees.* Washington, D.C.: National Institute of Justice.

Hser, Y., M.D. Anglin, T. Wickins, and J. Homer (1992). *Techniques for the estimation of illicit drug-use prevalence: An overview of relevant issues.* Washington: National Institute of Justice. (NCJ Research Report No. 133786.)

Hubbard, R., M. Marsden, and J. Rachel (1989). *Drug Abuse Treatment: A National Study of Effectiveness.* Chapel Hill: The University of North Carolina Press.

Jolly, G.M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* 52: 225-247.

Maddux, J. and D. Desmond (1986). "Relapse and recovery in substance abuse careers," in F. Tims and C. Leukefeld, (eds.) *Relapse and Recovery in Drug Abuse.* NIDA Monograph Series, 72: 49-72.

McLachlan, G.J. (1995). "Mixtures—models and applications," in N. Baladrishnan and Asit P. Basu (eds.) *The Exponential Distribution: Theory, Methods and Applications.* Amsterdam: Gordon and Breach Publishers.

McNagny, S.E. and R.M. Parker (1992). "High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic," *Journal of American Medical Association*, 267: 1107-1108.

Mieczkowski, T. (1990). "The accuracy of self-reported drug use: An evaluation and analysis of new data," in R. Weisheit (ed.) *Drugs, Crime and the Criminal Justice System.* Anderson Publishing.

Mieczkowski, T. (1992). "Immunochemical hair assays, urinalysis, self-reported use and the measurement of arrestee cocaine and marijuana exposure in a large sample," presented at the Annual Meetings, American Society of Criminology, New Orleans: November.

Nurco, D., T. Kinlock, T. Hanlon, and J. Ball (1988). "Nonnarcotic drug use over an addiction career-A study of heroin addicts in Baltimore and New York City," *Comprehensive Psychiatry*, 29: 450-459.

Seber, G.A.F. (1973). *The Estimation of Animal Abundance and Related Parameters.* London: Griffin.

Seber, G.A.F. (1965). "A note on the multiple-recapture census," *Biometrika*, 52: 249-259.

Simeone, R.S., W.T. Nottingham, and L. Holland (1993). "Estimating the size of a heroin using population: An examination of the use of treatment admissions data," *International Journal of the Addictions*, 27(2): 171-237.

Simeone, R.S., W.M. Rhodes, and D. Hunt (1995). "A plan for estimating the number of 'hardcore' drug users in the United States," *International Journal of the Addictions*, 30(6): 637-657.

Simeone, R. S., W.M. Rhodes, D. Hunt, and L. Truitt (1997). *A Plan for Estimating the Number of "Hardcore" Drug Users in the United States: Preliminary Findings*. Washington DC: Office of National Drug Control Policy.

Simpson, D. and K. Marsh (1986). "Relapse and recovery among opioid addicts 12 years after treatment," in F. Tims and C. Leukefeld (eds.) *Relapse and Recovery in Drug Abuse*. NIDA Monograph Series 72: 86-104.

Sobell, L., M. Sobell, and D. Riely (1988). "The reliability of alcohol abusers self-reports of drinking and life events which occurred in the distant past," *Journal of Studies on Alcohol*, 49: 225-232.

Toborg, M.A., J.P. Bellassai, A.M. Yezer, and R.P. Trost (1989). *Assessment of pretrial urine testing in the District of Columbia*. Washington, D.C.: National Institute of Justice (Grant No. 83-IJ-CX-K046).

Viveros, R. (1995). "Conditional inference and applications," in N. Balakrishnan and A.P. Basu (eds.) *The Exponential Distribution: Theory, Methods and Applications*. Amsterdam: Gordon and Breach Publishers.

Waldorf, D.(1973). *Careers in Dope*. Englewood Cliffs: Prentice Hall.

Wickens, T.D.(1993). "Quantitative methods for estimating the size of a drug-using population," *Journal of Drug Issues*, 23: 185-216.

Wish, E.D., J. O'Neil, and V. Baldau (1990). "Lost opportunity to combat AIDS: Drug abusers in the criminal justice system," in C.J. Leukefeld, R.J. Battjes, and Z. Amsel (eds.) *AIDS and Intravenous Drug Use: Future Directions for Community-Based Prevention*. NIDA Monograph Series 93.

Woodward, J.A., D.G. Bonett, and M.L. Brecht (1985). *Estimating the Size of a Heroin-abusing Population Using Multiple-recapture Census*. NIDA Monograph Series 57: 158-170.

Zacks, S. (1995). "System reliability and associated inference," in N. Baladrishnan and Asit P. Basu (eds.) *The Exponential Distribution: Theory, Methods and Applications*. Amsterdam: Gordon and Breach Publishers.